FACE POSE CORRECTION BASED ON DEPTH INFORMATION

BACKGROUND

Occasionally, when a user is using mobile device (e.g., a mobile phone, tablet, or the like) in a video call, the user may hold or place the mobile device at an angle relative to the face of the user such that it captures the user's image at an unsuitable angle, e.g., a less-than-optimal angle to view the user's expression, an angle that produces an image with facial distortion, etc. For example, a user may hold a mobile phone facing upwards in a way that the phone is pointed at an upward angle toward the user's face, which can capture the user at a less than optimal angle (e.g., looking upward from below the chin and nose) and may cause the face of the user to appear distorted or appear as if the user is not facing the other participants in the call and eye contact may seem lost. Additionally, the user's hand may be shaky and cause jitter in the image being captured for the video call.

The background description provided herein is for the purpose of generally presenting the context of the disclosure. Work of the presently named inventors, to the extent it is described in this background section, as well as aspects of the description that may not otherwise qualify as prior art at the time of filing, are neither expressly nor impliedly admitted as prior art against the present disclosure.

SUMMARY

Some implementations are generally related to mobile device video calling, and in particular to pose correction of video images captured by a mobile device.

Some implementations include capturing a user's face pose and an eye gaze using a depth camera associated with a user device and performing face pose correction and stabilization of the face pose.

Some implementations can include a computer-implemented method. The method can include receiving a video including a plurality of frames, the video captured by a physical camera at a first point of view, wherein the video includes depth information corresponding to the plurality of frames, and detecting a face within the video, wherein the face is within a foreground portion of one or more frames of the video.

The method can also include determining the foreground portion of the plurality of frames based on one or more depth values of the depth information corresponding to the plurality of frames, and positioning a virtual camera at a second point of view, wherein the second point of view is different from the first point of view. The method can further include obtaining a projection matrix of the foreground portion based on the virtual camera, the projection matrix corresponding to the second point of view, and generating a modified video that includes a modified foreground portion based on the projection matrix.

The method can also include adjusting the projection matrix of the foreground portion to reduce a lens effect of the physical camera used to capture the video. In some implementations, detecting the face can include extracting the face out of a background plane, wherein the background plane is determined based on the depth information. The method can also include rotating the background plane to face the virtual camera at the second point of view.

In some implementations, determining the foreground portion can include detecting the face and a background, and the detecting can include extruding the foreground portion of the face from the background. In some implementations, extruding can include can include detecting face depth based on the depth information, determining a depth within the frames of an outer edge of the face, and setting the background to include portions of the frames at a depth of the outer edge of the face and at depths beyond the outer edge of the face.

The method can also include comprising blurring the background. The method can further include determining a temporal history of depth information of the video. The method can also include comprising providing hysteresis of video re-projection using the temporal history of the depth information of the video so as to maintain continuity of a re-projection and not respond to temporary objects moving in front of the face being re-projected, wherein the temporary objects are in front of the face being re-projected for less than a threshold period of time.

Some implementations can include a system comprising one or more processors coupled to a non-transitory computer readable medium having stored thereon software instructions that, when executed by the one or more processors, cause the one or more processors to perform operations. The operations can include receiving a video including a plurality of frames, the video captured by a physical camera at a first point of view, wherein the video includes depth information corresponding to the plurality of frames, and detecting a face within the video, wherein the face is within a foreground portion of one or more frames of the video.

The operations can also include determining the foreground portion of the plurality of frames based on one or more depth values of the depth information corresponding to the plurality of frames, and positioning a virtual camera at a second point of view, wherein the second point of view is different from the first point of view. The operations can further include obtaining a projection matrix of the foreground portion based on the virtual camera, the projection matrix corresponding to the second point of view, and generating a modified video that includes a modified foreground portion based on the projection matrix. The operations can also include stabilizing the modified foreground portion of the modified video.

The operations can further include adjusting the projection matrix of the foreground portion to reduce a lens effect of the physical camera used to capture the video. In some implementations, detecting the face comprises extracting the face out of a background plane, wherein the background plane is determined based on the depth information.

The operations can also include rotating the background plane to face the virtual camera at the second point of view. In some implementations, determining the foreground portion includes detecting the face and a background, where the detecting can include extruding the foreground portion of the face from the background. In some implementations, extruding can include detecting face depth based on the depth information, determining a depth within the frames of an outer edge of the face, and setting the background to include portions of the frames at a depth of the outer edge of the face and at depths beyond the outer edge of the face.

The operations can also include blurring the background. The operations can further include determining a temporal history of depth information of the video. The operations can also include providing hysteresis of video re-projection using the temporal history of the depth information of the video so as to maintain continuity of a re-projection and not respond to temporary objects moving in front of the face